

On-line Access to Archives (and Other Records) in the Digital Age Australian Society of Archivists Annual Conference, Parramatta (October, 2016)

I'd like to talk about two aspects of on-line access : the Big Picture and Federated Access. This is not about how to choose your next project or how to undertake it. It is not about how any one archives programme decides what to digitise in the medium term. This presentation will touch lightly on how it is all to be paid for and how that expenditure nation-wide (mostly consisting of taxpayer dollars), should be managed and directed. I want to focus, however, on some policy and strategic considerations – how to conquer Russia, not how to capture Smolensk or get out of bed.

In April, this year, it was reported in [itnews](#) that SLNSW was going to ask the private sector to take over its digitisation programme in the hope that in so doing, they could “make a buck” out of it. I posted this on the archives-and-records listserv and some of the issues that emerged in [the ensuing debate](#) are here displayed on screen. The State Library, it was reported, was nearing the end of a 10-year \$72m “journey” that had resulted in the digitisation of less than 1% of its total “collection”.

| Big Picture | |
|--|---|
| Napoleon in 1812 | Digitisation in Australia |
| Get out of bed ↳ Capture Smolensk ↳ Conquer Russia | My next project ↳ Our 5-year priorities ↳ National strategy |

| Mass Digitisation |
|--|
| <ul style="list-style-type: none"> • Immense benefits from online searching of digitised & born digital resources with further potential for re-presentation by third parties. • How much of the archival heritage can realistically be digitised any time soon? • Can online searching be integrated for both - <ul style="list-style-type: none"> ♦ digitised and un-digitised resources? ♦ gathered and ungathered resources? • Does access to our stuff have requirements specific to our class of material? <ul style="list-style-type: none"> ♦ Does online access integrated with other resources vitiate them? ♦ What are our requirements and how far are they being met? ♦ Can we reconceptualise description to satisfy them? • Doubts & Queries - <ul style="list-style-type: none"> ♦ Is content vs contextualised searching dumbing down archival research? ♦ Does digitisation (facilitating access to some but not all resources) distort research capabilities & privilege some sources and their users? |

If \$72m buys you digitisation of “less than 1%” it would appear that it would cost \$7.2b to digitise it all. We don't know if it is 1% of everything or just 1% of that portion that SLNSW thinks worthy of digitisation. Overseas figures suggest that such metrics (the % of digitisation already done) are often calculated on the basis of what those reporting the figures think is worthy of digitisation – excluding large volumes of heritage resources that are not. Nor do we know by what quantities the SLNSW collection is growing each year from new deposits of un-digitised material. So the total cost of digitising everything after a further 25 years (say) of growth could be considerably higher than \$7.2b.

When we compare the relatively puny holdings of libraries and manuscripts collections with the massive runs in government and corporate archives, the total figure for documentary heritage resources must be huge. But let's be generous and assume that SLNSW accounts for 2.5% of the total of the total national documentary heritage. This means that, without taking into account any future growth rate in the total quantity of un-digitised resources held by these institutions, the bill for digitising the existing national heritage in its entirety would be \$288b. That leaves out of consideration all the ungathered records of heritage value that do not ever end up in a library or archives. Shortly afterwards, a report on the situation in Finland appeared on-line suggesting a similar order of magnitude there.



What kind of thinking sits behind all this? If there is any strategic thinking setting all this out, I would like to see it. You don't get many laughs at my age. Are people just mindlessly digitising because it's a good thing to do? Is any thought being given to what we are trying to achieve nationally? What are the national policy and strategic goals? Are there any or are we just doing as much as we can as fast as we can? Is it realistically supposed, as some of the reporting suggests, that if we just put our heads down and keep on scanning sooner or later we will reach the end of a very, very long road? One thing is very clear: for the foreseeable future we'll have to integrate on-line access to digitised, born digital, and non-digitised resources.

As many of our archives and libraries already do, users will go on being presented with results screens, which often still take the form of interminable item inventories (so 20th century!), displaying icons showing which have been digitised enabling to user to reach the digitised resource in one click and leaving them to make do with the description only for the rest. So, for a long time to come, on-line access will have to be provided for both digitised resources and, probably a lot more, un-digitised resources (to say nothing of born-digital resources). Leaving aside digitisation for preservation (which I don't think involves more than a fraction of the

total) and the advantages of on-line distribution of the asset (which I accept is a compelling motive for mass digitisation), what are the issues? Here are some.

| Public/Private Partnerships | Strategic implications |
|---|---|
| <ul style="list-style-type: none"> • Proliferation? Will there be a range of suppliers that are diverse, non-standardised, and potentially erratic? • Standardisation? Are there standardised contracts (or statements of minimum requirements) or is each institution sovereign and free to make its own arrangements? Is access free? Are charge rates standardised? Do they last forever? Do they lock out free services such as TROVE? • National interest? Is this wise policy? Who has oversight of the heritage policy and public interest aspects? • Mandate? Should limits be placed on sovereign cultural institutions in the digital exploitation of the resources they hold? They were set up in the pre-digital era. Is it time to re-examine their mandates? Do the assumptions behind their establishment support moving to digital processes designed to “turn a buck”? | <ul style="list-style-type: none"> • Oversight: How is this large and growing expenditure of public money being planned and monitored? Are priorities assigned across institutions and jurisdictions? Are we digitising what needs doing or what sits in the vaults of those who are best at selling? • Preservation: How to balance access and preservation? Do we stop preserving digitised resources and move conservation spending onto those which aren’t? Do we stop physical preservation altogether, close down conservation facilities, sack conservators, and move our spending into digital preservation? • Generic tools: Should our stuff be accessed the same as other heritage stuff? Does ours have unique characteristics? Does content searching (OCR) mean the loss or diminution of contextual understanding? Does using generic tools vitiate our responsibility to curate and provide access in more appropriate ways? • Description: Should we pay more attention to descriptive practices that produce on-line descriptions to improve and better integrate on-line representations of resources irrespective of whether they have been digitised or not? What are our requirements? Are we just flying by the seats of our collective pants? |

Firstly, as to the PPP proposal :

- Does it mean a proliferation of across the sector with a range of suppliers that are diverse, non-standardised, and potentially erratic? Is access free? Are charge rates standardised and do they last forever? Do they lock out free services such as TROVE?
- Even if they are commercially advantageous, is this wise policy from a national (as distinct from a jurisdictional or institutional) point of view?
- Are there even standardised contracts (or statements of minimum requirements) when these deals are entered into, or is each institution sovereign and free to make its own arrangements?
- Who is over-seeing them? I’m not talking about Treasury oversight of the contracts but oversight of the heritage policy and public interest aspects.
- What limits (if any) should be placed on sovereign cultural institutions in regard to the digital exploitation of the resources they are entrusted with?
- Most of these institutions were set up in the pre-digital era? Does their mandate cover this or is it now time to re-examine their mandates? Is this a case of function-shift?
- Should the assumptions behind their establishment as custodial bodies remain unscrutinised as they move towards digital processes designed to “turn a buck”?

Secondly, as to the strategic implications for cultural heritage policy :

- Mass digitisation, on any reckoning, represents a large and growing expenditure of resources (mostly taxpayers’ money). How is this public spending being planned and monitored? Is expenditure across institutions and jurisdictions being evaluated to assign priorities or is each sovereign institution free to compete with all the others without any guiding hand attempting to prioritise? Are we digitising what needs to be digitised or what sits in the vaults of those who are best at selling their case?
- How do we prioritise digitisation for access and digitisation for preservation? Do we stop preserving resources that have been digitised and move conservation spending onto those which haven’t? Do we stop spending money on physical preservation altogether, close down the conservation facilities, sack the conservators, and move all our spending into digital preservation?
- Should our resources be accessed in the same way as other heritage resources? Do archives and records have unique characteristics that mean we cannot simply emulate generic deployment and search methodologies? If we emulate such strategies are we vitiating our responsibility to curate and provide access in ways more appropriate to the materials we manage? Specifically, does content searching (OCR) mean the loss or diminution of contextual understanding?
- If the majority of resources remain un-digitised should we be paying a lot more attention to descriptive practices that produce on-line descriptions? What is being done to improve and better integrate the on-line representation of resources irrespective of whether they have been digitised or not? What are we doing to provide federated searching across institutional and jurisdictional boundaries? Have we defined our requirements for federated searching or are we just flying by the seats of our pants?

- Is mass digitisation distorting the research process itself by providing uneven access opportunities to digitised over non-digitised material, privileging users of digitised resources over other users, and unevenly portraying the stories of some of the subjects of records over others?

One response often given to such questions is that they don't matter. On-line access to digitised resources is incredibly popular and gives access in ways that are simply impossible otherwise to millions of potential users who would not otherwise use them – and no one can argue with that. Examining the inadequacies of content searching does not mean, however, that we want to prevent it and deprive anyone of its advantages.

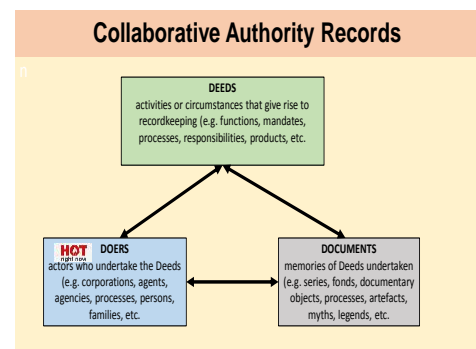


Exploring critically the imbalance resulting from privileging some resources and users over others does not mean we want to stop digitisation altogether. This is not about denying the advantages of digitisation, it is about keeping our eyes open to its shortcomings. It's about saying that an intelligent assessment goes beyond undiluted triumphalism and onto a sober consideration of how to make things better.

Two years ago, in Christchurch, I made a [Modest Proposal](#) for improving things. For me, this has a long history. In

February, 1971, on my very first day in what is now NAA, Peter Scott placed into my hands two loose-leaf binders called Handbooks – one for the Commonwealth and another for Australasia and told me that my job was, inter alia, to fill them up. Peter intended them to display material that would contextualise the heritage resources of the Commonwealth Government and of the nation at the highest (bird's eye) level. It was a task I never completed (or even properly started) but it haunted me for 15 years until the (now defunct) Council on Archives asked me to do a [report on standardisation](#).

My 1987 Report preceded most of the descriptive standards efforts of the last 30 years – I think Canada, the US, and Michael Cook in the UK may have been just in the early stages of their work. It certainly preceded encoding techniques and mark-up strategies that led to EAD, EAC, and the like. But the methodology involved in those efforts, getting everyone to describe material in the same way or evolving methods to normalise non-standardised descriptions so they would appear similar, was already foreseeable. I decided to recommend another way – one that I hoped Peter would approve of - what I now call the road not taken. It involved what is essentially a registration process – leaving everyone to describe things in their own, non-standardised way and to integrate by linking or reporting the results into a single, unified, shared, collaborative framework that would bring them together. The example I use to illustrate the need for this is the Superintendent's records from the District of Port Phillip when it was part of NSW. Deposits are described in both the Archives Office of NSW and in the PROV. The framework (or national register) would bring them together without requiring either to modify their descriptive practices and would provide a universal view of the totality of archival heritage resources at a bird's-eye level that could be comprehended and navigated.



In 1987, the purpose of descriptive standardisation was usually seen as drawing together records of common provenance.

Nowadays, we understand that it spans the entire range of recordkeeping activities¹, viz.

- **Deeds:** activities or circumstances that give rise to recordkeeping (functions, mandates, processes, responsibilities, products, etc.);
- **Doers:** actors who undertake the Deeds (e.g. corporations, agents, agencies, processes, persons, families, etc.);

¹ I have included "process" in all of them to illustrate that the same thing can be described as an instance of two, or even all three, entity-types.

- **Documents:** memories of Deeds undertaken (e.g. series, *fonds*, documentary objects, processes, artefacts, legends, etc.).

In the intervening years, the road that was taken has seen, inter alia, the development of descriptive standards, metadata standards, and encoding systems - all focussed on how we should go about the formation of actual descriptions in order to bring them into some kind of alignment. Recently, I am pleased to say, efforts have been made to take the other path – not in opposition but in fulfilment of this process. These contextualising frameworks from France, Germany, and the US still emphasise Doers, providing authority data for shared use – but it’s a start.

| Collaborative Authority : EAC-CPF | Collaborative Authority: SNAC (social networks & archival context) |
|--|--|
| <ul style="list-style-type: none"> • collaborative project : Association of French archivists & Archives of France (cf ICA 2012: Claire Sibelle). Similar in USA and Germany, etc. • to create standardised descriptions of records creators (or to share the existing information) and propose patterns archival institutions could reuse to describe their own records creators • to provide French Departmental Archives with an easy-to-use, interactive and participatory tool • ISAAR(CPF) is the reference content standard and EAC-CPF is the communication standard • re-usable and real exchangeable authority records which could be imported in their own information system | <ul style="list-style-type: none"> • collaboration among NARA (US), Institute for Advanced Technology in the Humanities (U. of Virginia), California Digital Library, & School for Information Science at U. of California, Berkeley. • separates descriptions of persons, families, & organizations from description of historical resources • provides integrated access to material held by archives and libraries, large and small + setting the stage for cooperatively maintaining information about people documented in collections • an international cooperative to enable archivists, librarians, scholars, and eventually “citizen archivists” to maintain and add biographical-historical data • prototype draws on more than 2.6 million (soon to grow to more than 3.5 million) descriptions of persons, organizations, and families |

Four years after my Report to the ACA, we released the PROV’s first *Summary Guide* – the results of years of effort to upload our descriptions into computerised form and produce a comprehensive machine-generated finding aid (I think the first in Australia). Towards the end of that project, I decided this was not good enough so, partly to illustrate what the Victorian contribution to a national register would look like, and partly to put in a keystone missing from the arch that spanned the comprehensive but very detailed and lengthy catalogue of Victoria’s public records, I wrote a 100 page [Digest of the Public Records of Victoria](#) to go with the *Summary Guide* linking Victoria’s public records to the history of the State and briefly summarising the scope and functions of the 100 or so Record Groups into which our descriptions were divided. Our purview included the entirety of Victoria’s public records – not just what we had in custody. Since then, federated searching has evolved using technological developments predictable, but not then available.

It is still my conviction that archival materials have special requirements for curating and preserving contextual meaning by description of both breadth and depth that cannot be met by simply adapting generic federated search methodologies applied to other heritage resources. For some years now I have been promoting the *Modest Proposal* as a form of federated access based on registration. This is an “as well as” not an “instead of”. It is an attempt to put aside the many efforts now up and running that provide actual federated searching and asking what our functional requirements are.

| Our requirements |
|---|
| <ul style="list-style-type: none"> • Depth • Protean • Inclusive • Wholistic • Collaborative • Authentic • Differentiated description • Differential access |

What are we (or should we be) trying to do? When we know that we can evaluate the many efforts and maybe improve them or replace them with something else. This is not founded on any expectation that when we have enunciated our requirements they will be what our users are crying out for. We know what they want : they want access to the stuff, immediately, comprehensively, and without complication. A National Register doesn’t give them that. But our job is more than giving them what they ask for. Always has been. We are not just in the user-satisfaction business. We are also in the preservation business – preserving the contextual meaning of the materials we handle so that, beyond being displayed for widespread and

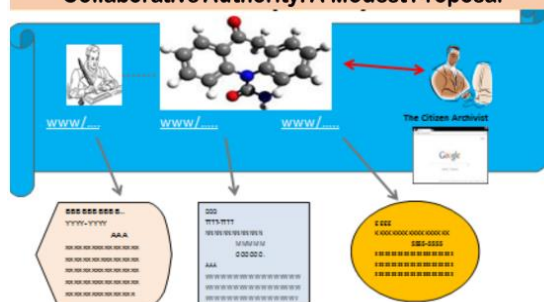
convenient use, they continue to have the authority and integrity of the evidential materials they were to begin with and must continue to be while they remain in our care².

Archives Portals: through the Looking-Glass



world in which her new reality is bent into shape by activity and circumstance attendant on its formation. That world is made up of our descriptive endeavour to portray, as accurately as we can, the complex contextual structures that link objects to documentary, organisational, familial, and functional processes in an ever-widening framework of relationships that impose upon her understanding the meaning and truth of what she finds as well as supporting her on-going journeys of discovery.

Collaborative Authority: A Modest Proposal



A Modest Proposal : The Framework

TRAUNCHE ONE : CATEGORIES based on ANZSIC work classification

- A. AGRICULTURE, FORESTRY, & FISHING AAA 003 (--- Agriculture; --- Horticulture; etc.)
- B. MINING AAA 004 (--- Coal; --- Ores; --- Liquid Minerals; etc.)
- C. MANUFACTURING AAA 093 (--- Plants, Factories, Mills; --- Small Business & Home Products) etc. etc.

TRAUNCHE TWO : TOPICS

- A. ABORIGINES & TORRES STRAIT ISLANDERS AAA 195 (--- Conflict (with Aborigines); --- Stolen Generations; etc.)
- B. IMMIGRATION AAA 045 (--- Asylum-Seekers & Refugees; --- Colonisation & Settlement; etc.)
- C. LAND SETTLEMENT AAA 085 (--- Agricultural & Pastoral Leases; --- Closer Settlement; etc.) etc. etc.

TRAUNCHE THREE : JURISDICTION

- A. BRITISH EMPIRE & COMMONWEALTH AAA 046 (--- Kingdom of Great Britain; --- Kingdom of Great Britain & Ireland; etc.)
- B. AUSTRALIA AAA 054 (--- Government of NSW; --- Government of Victoria; etc.)
- C. NEW ZEALAND AAA 064 etc. etc.

TRAUNCHE FOUR : LOCATIONS from ASGC classification on ABS web site

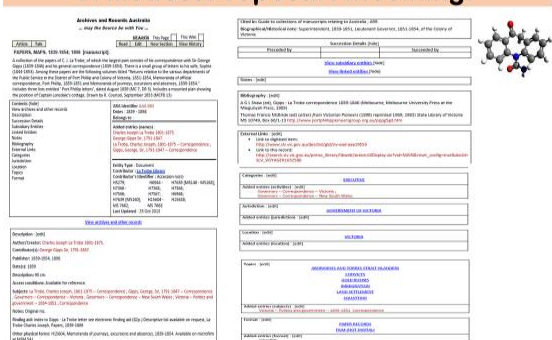
- A. AUSTRALIA AAA 097
- NSW AAA 098 (--- Regions & Districts; --- Local Government Areas; --- Cities & Major Towns)
- VIC AAA 099 (--- Regions & Districts; --- Local Government Areas; --- Cities & Major Towns) etc. etc.

TRAUNCHE FIVE : DOCUMENT TYPE/FORMAT

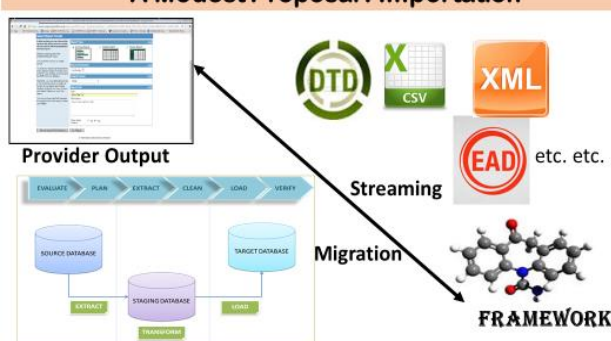
- A. PAPER RECORD AAA 981 (--- Documents; --- Volumes; --- Cards; etc.)
- B. DIGITAL RECORDS AAA 430 (--- Image; --- Data; etc.) etc. etc.

The basic idea was to establish a collaborative framework for Australian descriptive endeavour that would satisfy most (if not all) the functional requirements identified in my proposal. It would enable on-line descriptions (at any level) to be linked to one or more of five categorisations – some of which would only involve borrowing or adapting classifications already maintained by other bodies such as the Bureau of Statistics. To give a concrete example of how this might work, I developed a model of a Wiki into which technologically challenged contributors (those I call the barefoot archivists) could submit and the more advanced contributors might steam their descriptive efforts. Many other implementation strategies might be

A Modest Proposal : Incoming!



A Modest Proposal : Importation



adopted. I use the wiki merely to illustrate the idea – not to put forward a project proposal. The technology for federating resources is improving all the time. It certainly goes far beyond what was possible in 1987 when the basic idea was first put forward.

² I will not recapitulate the requirements at length here. They are set out in my presentation to the ASA/ARANZ Joint Conference in Christchurch (2014) and that presentation is available on my web site.

A Modest Proposal: Results Page

Your search for **C J La Trobe** has produced the following results. Filter results by Date ▼

DOCUMENTS (18,567)
Documents, images, film, video, maps, plans, audiovisual and other records described singly, in series, or in other groupings.

→ Filter Documents by Format ▼

PERSONS (503)
Responsibilities assigned to those who produce the records.

CATEGORIES (250)
Rules, activities, and jobs of those who produce or are the subject of the records.


TOPICS (12)
Subjects of the records.

PLACES (89)
Persons and corporations who produce the records or are dealt with.

CONTRIBUTORS (16)
Persons and corporations who hold or report on the records.

JURISDICTIONS (3)
Federal, state, and local government administrations who produce the records.

LOCATIONS (130)
Places where the records are produced or to which they relate.



Last year, in Hobart, Michael Piggott led a team session proposing that action be taken by our community to present for public consumption a survey and evaluation of what has been and what is being done to document Australian society. This would not be a listing of available resources but a commentary on the adequacy and shortcomings of recordkeeping efforts, past and present. I was much impressed when someone from the

Documenting Australian Society

TRAUNCHE ONE : CATEGORIES

: 03 INTERNET [★] **hot topic**

1K. FINANCIAL & INSURANCE SERVICES

History :

In the earliest period, the NSW economy was without currency or financial institutions. Government control over housing, stores and other goods and services was complete. With emancipation, the arrival of free settlers, alienation of Crown land, and the rudiments of commerce a system of barter, promissory notes, and reliance on foreign coinage became the norm...

Documentation:

This sector has seen frequent collapses, mergers, and transformations. The larger institutions have been relatively stable since the last mopping up of the smaller banks and amalgamations of the bigger investment and life businesses by the big banks. Most of the banks (including the Reserve) now have archives programmes, some of which include legacy records resulting from earlier mergers...

Accessibility:

Banks' customer records are subject to strict limits on accessibility founded in law on *Tournier's Case* and in various statutes enjoining strict confidentiality. The situation is similar with other financial services but the boundary line between the confidentiality of financial services records and that of ordinary commercial transactions is not always clear...

Needs:

The arrangements are very patchy. Protocols for the disposition of records during liquidation or merger derive more from business than archiving principles...

: 01 BANKING



© Chris Hurley 2016